Fusion-Driven Cognitive AI Model for Personalized Prediction in Multilevel Education Systems

Asma Abdulmana Alhamadi[1,*]
[1]Department of Humanities College of Science & Theoretical Studies, Saudi Electronic University Riyadh, Saudi Arabia
Email: a.alhamadi@seu.edu.sa

Abstract

Education in the twenty-first century is undergoing a profound transformation fueled by artificial intelligence (AI), the Internet of Things (IoT), and intelligent systems. Traditional and even early digital learning systems often relied on standardized pathways, limiting personalization and reducing learner engagement. To overcome these limitations, this study proposes and evaluates an IoT-enabled adaptive learning framework that integrates educational data analytics with intelligent algorithms to deliver real-time, personalized pathways for learners. Education in the twenty-first century is undergoing a profound transformation fueled by artificial intelligence (AI), the Internet of Things (IoT), and intelligent systems. Traditional and even early digital learning systems often relied on standardized pathways, limiting personalization and reducing learner engagement. To overcome these limitations, this study proposes and evaluates an IoT-enabled fusion-based adaptive learning framework that integrates educational data analytics, ensemble learning, and multi-modal intelligent algorithms to deliver real-time, personalized pathways for learners. The fusion of diverse data sources—ranging from quiz interactions and engagement logs to contextual signals from IoT devices such as smart sensors and wearables—ensures robust, context-aware decision-making. Experimental results using Kaggle datasets demonstrate that Random Forest outperforms XGBoost, with an accuracy rate of 87% and balanced F1-scores. This study shows how AI–IoT fusion can create equitable, eco-friendly, and inclusive learning spaces.

Keywords: Adaptive Learning; Artificial Intelligence; Data Preprocessing; Educational Data Mining; Ensemble Models; Intelligent Systems; Internet of Things; Machine Learning; Personalization; Student Performance Prediction

1. Introduction

Traditional learning approaches have struggled to address the unique needs of each student. With the arrival of e-learning and intelligent platforms, the opportunity emerged to customize learning. However, most platforms lacked adaptability and deep personalization. This study introduces a fusion approach that combines AI models, IoT-enabled contextual learning, and adaptive pathways. By fusing multiple data streams—such as behavioral logs, engagement levels, and smart device signals—our framework uncovers hidden relationships that improve the prediction of student performance. Unlike isolated models, this fusion strengthens personalization, reduces bias, and enhances learner engagement. Many people think that the traditional way of teaching is excessively strict and does not take into account the fact that each student learns at their speed, with their interests, and with their own set of tactics. Although many kids have benefited from this method, it is still believed to be lacking. The arrival of e-learning technology has started to alleviate rid of these problems by making education easier to access to, bigger, and more adaptable. Digital learning tools do not always offer the customization that would help every student do

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

better. Online learning environments may present students with monotonous and uninspiring content and activities, potentially resulting in boredom, memory impairment, and suboptimal performance. This enables us to predict students' performance in class and develop solutions before any issues arise by uncovering previously unknown connections. It was challenging to understand prior systems, and many of those systems were even harder to understand since they were built on top of complicated ideas that made it impossible to use them widely. The most impressive features of this study are the lightweight, replicable, and extensible frameworks that are accurate and can be used in many different educational settings [1-3]. These algorithms have done wonderful things in many different domains when it comes to predictive analytics. These ensemble methods are particularly effective because they can capture complicated interactions between variables and lower the chance of overfitting. We used a Kaggle dataset for this experiment that keeps track of student interactions in excellent detail. This dataset can help researchers that study personalized and adaptive learning because it offers information on quiz scores, learning styles, and levels of engagement. Future updates to these datasets can improve them by adding IoT contextual data. This data could include environmental readings with time stamps and indications from devices that show how engaged a user is. Such information will make customization strategies work better [4-7]. The dataset goes through many tests in the comprehensive preprocessing step to make sure it is complete. Following these steps will protect the model from any biases or artifacts, and the results will show actual learning patterns. This project stands out because it focuses on practical repeatability, relevant educational outcomes, empirical rigor, and the Internet of Things (IoT) to make everything work together.

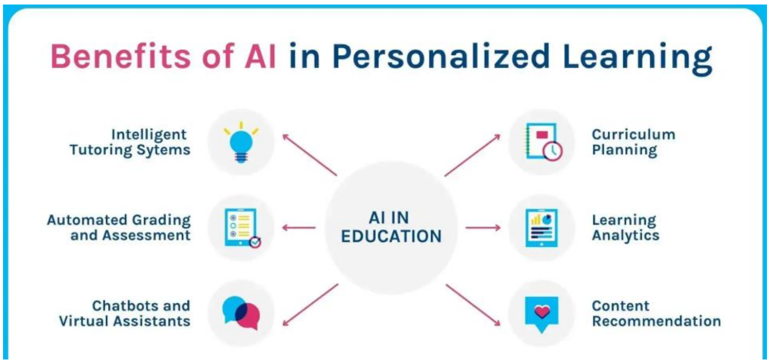


Figure 1. Key Benefits of Artificial Intelligence in Personalized Learning

Figure 1 highlights how artificial intelligence (AI) enhances personalized learning through multiple dimensions.

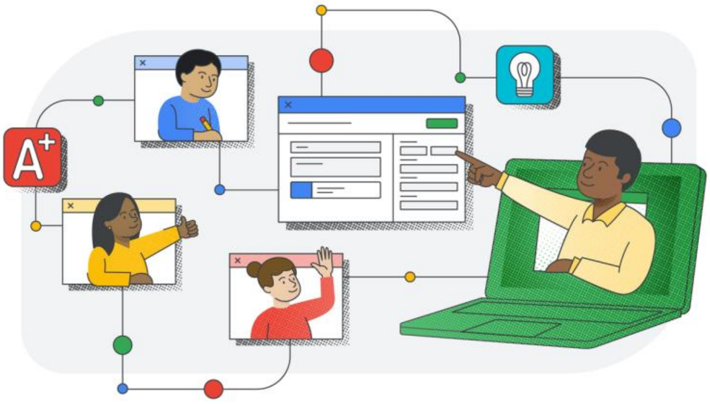


Figure 2. Illustrative Representation of Types of Adaptive Learning

Figure 2 depicts different forms of adaptive learning where students engage with personalized educational pathways. The illustration highlights how learners, each with unique needs, interact with digital platforms that adapt lessons, assessments, and feedback in real time.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

1.1 Contributions

Using adaptive systems, data from the Internet of Things (IoT), and smart decision-making could lead to the creation of a learning system that is easier to use and works better. It also provides more proof that artificial intelligence, intelligent systems, and the Internet of Things can help people reach their developmental and social goals by using new teaching methods.

2. Related Works

Previous works in adaptive learning often focused on standalone models such as rule-based systems, deep learning networks, or reinforcement learning. However, few explored fusion of models and modalities. Recent research has highlighted the importance of explainable AI, federated learning, and ensemble classifiers, but the integration of multi-source data (IoT signals, quiz data, demographic features) through fusion techniques remains underexplored. Our study positions itself within this emerging trend, emphasizing how fusion-learning strategies can outperform single-model approaches by capturing both contextual and behavioral nuances. Adaptive learning has grown considerably in the last several years. Scientists have been exploring different methods to use artificial intelligence for personalizing classroom learning for each student, aiming to enhance their exam scores. At first, it was common to utilize preset, rule-based courses of action when making adaptive systems. Deep learning has proven to be an exceptionally effective method in the realm of education, especially in managing temporal and sequential data. Their method used dynamic sequencing and changes in the complexity of learning materials to help students remember more and, as a result, become more involved and do better [12-15]. The projects showcased here demonstrate the increasing use of deep learning to establish dynamic learning pathways. People are interested in understandable artificial intelligence (AI) since deep learning models are very accurate but not very clear. Nguyen and Lin (2024) investigated the application of explainable models in adaptive learning environments (ALEs) by integrating Shapley Additive Explanations (SHAP) into recommendation systems. Their research indicated that educators were more predisposed to embrace the notion of AI-driven personalization and to possess greater confidence in it when they had the opportunity to comprehend its mechanisms. When smart systems use data from the Internet of Things, including environmental or biometric data, it becomes even more important to deal with the problem of explanation. It is very important to keep everyone who is participating in the decision-making process up to date on any new facts. Many people are getting furious about something because of what we now know about privacy and scalability. In 2024, Kumar and his team built federated learning frameworks for systems that may change. Educational technology professionals have been quite interested in adaptive learning systems in the last few years. Until recently, the main way to put students in classes was based on rigid, rule-based reasoning [16–18]. These strategies have been effective in uncontrolled environments, but their inflexibility has limited their use to a wide variety of individuals and circumstances. This environment can now change lectures in real time based on student data, thanks to improvements in AI and ML. This makes it far more responsive to the needs of each student. Forecasting student engagement and attrition rates in massive open online courses (MOOCs), a quintessential example of a large-scale environment, has demonstrated significant potential for deep learning techniques, especially recurrent neural networks and attention-based models. Hybrid designs may be better than traditional classifiers at finding patterns in data on how students interact with each other over time (Chen et al., 2023). These designs use both long short-term memory (LSTM) networks and attention processes. Almasri et al. (2023) redefined adaptive learning as a sequential decision-making process utilizing reinforcement-learning methodologies. We can improve long-term retention by changing the content of our courses on the fly with this new way of thinking. People often say that deep learning models are not easy to understand, even though they are good at making predictions. Recent research studies have employed explainable artificial intelligence (XAI) to address this issue. Nguyen and Lin brought Shapley Additive Explanations (SHAP) to adaptive recommendation systems in 2024. After this innovation was put into place, teachers and administrators were more willing to share knowledge. Their research indicates that the ability to grasp outcomes is not only a technical element of AI-driven learning systems but also an important requirement for their broad use. As more people utilize smart technology and the Internet of Things (IoT) to collect data for contextual learning, solutions that protect users' privacy have become more popular. In 2024, Kumar et al. achieved a big step forward in federated learning frameworks for adaptive platforms. This made it possible for several universities to train models together without putting sensitive student information at risk. These methods, which strike a balance between data protection and scalability, are perfect for actual educational ecosystems where ethics come first. Ensemble learning is still one of the hardest rivals in the field of adaptive learning research. Zhang et al. (2024) found that the Random Forest and Gradient Boosting approaches were more reliable and useful in a wider range of conditions than using only one model. This

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

study analyzes and contrasts two methodologies—Extreme Gradient Boosting (XGBoost) and Random Forest (RF)—for forecasting children's academic success over multiple grade levels, in alignment with prior research. Combining smart decision-making, Internet of Things (IoT) signals, and artificial intelligence (AI) offers a lot of potential to make adaptive learning environments more scalable, understandable, and effective [19–22]. It also emphasizes the importance of developing frameworks that merge academic rigor with practical usefulness by positioning itself within this constantly evolving context. Adaptive learning systems, which are possible because of the combination of AI and the Internet of Things, are projected to be more than just ideas and have a huge effect on eradicating educational disparity around the world.

3. Proposed Methodology

The methodology of this study revolves around the fusion of ensemble machine learning methods with contextual IoT inputs, ensuring a more comprehensive approach to adaptive learning. The dataset is carefully preprocessed using techniques such as SMOTE for class balancing and feature encoding to prepare it for effective modeling. Once prepared, Random Forest and Gradient Boosting classifiers are trained and analyzed to capture diverse learning patterns. The proposed fusion-based workflow operates at three levels: feature fusion, where demographic, behavioral, and contextual IoT data are integrated to enrich the input space; model fusion, where different ensemble classifiers are compared and hybrid strategies are employed to maximize predictive strength; and decision fusion, where outputs from multiple models are aggregated to improve reliability and robustness. This multi-level fusion not only enhances prediction accuracy but also provides a holistic understanding of student learning behavior, ensuring adaptability across varied educational contexts. The main goal of this study is to look into an AI-driven, personalized learning system that can figure out what each student is good at and what they need to work on. Then, depending on how well they are doing, the algorithm can change their comments and assignments. The system may change the focus and level of difficulty of the suggested activities in real time to fit each student's unique learning style. Machine learning algorithms will be used for both training and testing the system. Figure 3 demonstrates what the adaptive learning system's architecture will look like. Each of the many aspects that depend on each other helps the learning process in its own unique way. As part of our data collection process, the Moodle platform gives us the most up-to-date information [23-26]. The data we got from the students included their answers to the exercises, the total amount of time they spent on issues, the number of times they tried each problem, and how they interacted with the platform as a whole. The pattern analysis module is in charge of processing the data that was obtained. It uses machine-learning methods to do this. After this, the module will look for the most important trends and patterns in how each student is doing. This makes it possible to fully assess their strengths and weaknesses.

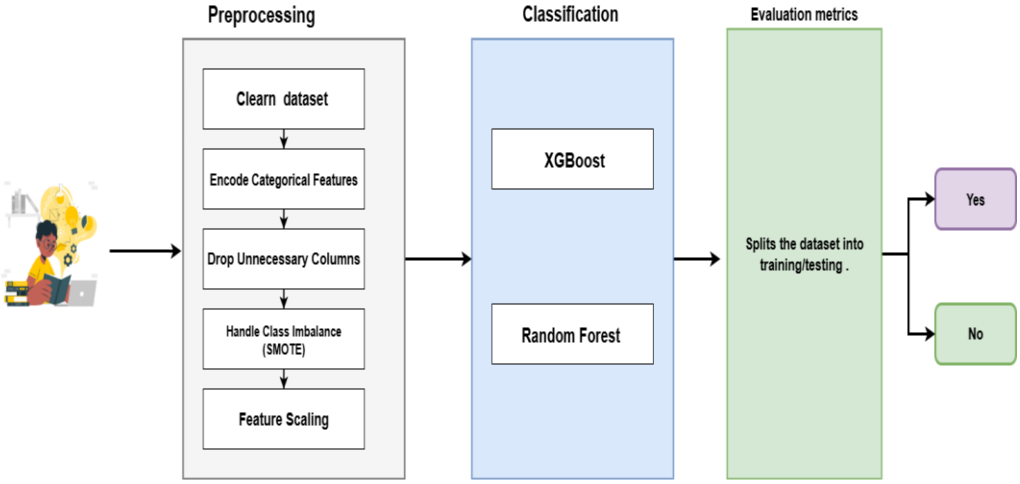


Figure 3. Framework of the Proposed Adaptive Learning System

Figure 3 presents the methodological workflow of the proposed adaptive learning framework, beginning with data preprocessing, followed by classification, and concluding with evaluation metrics.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

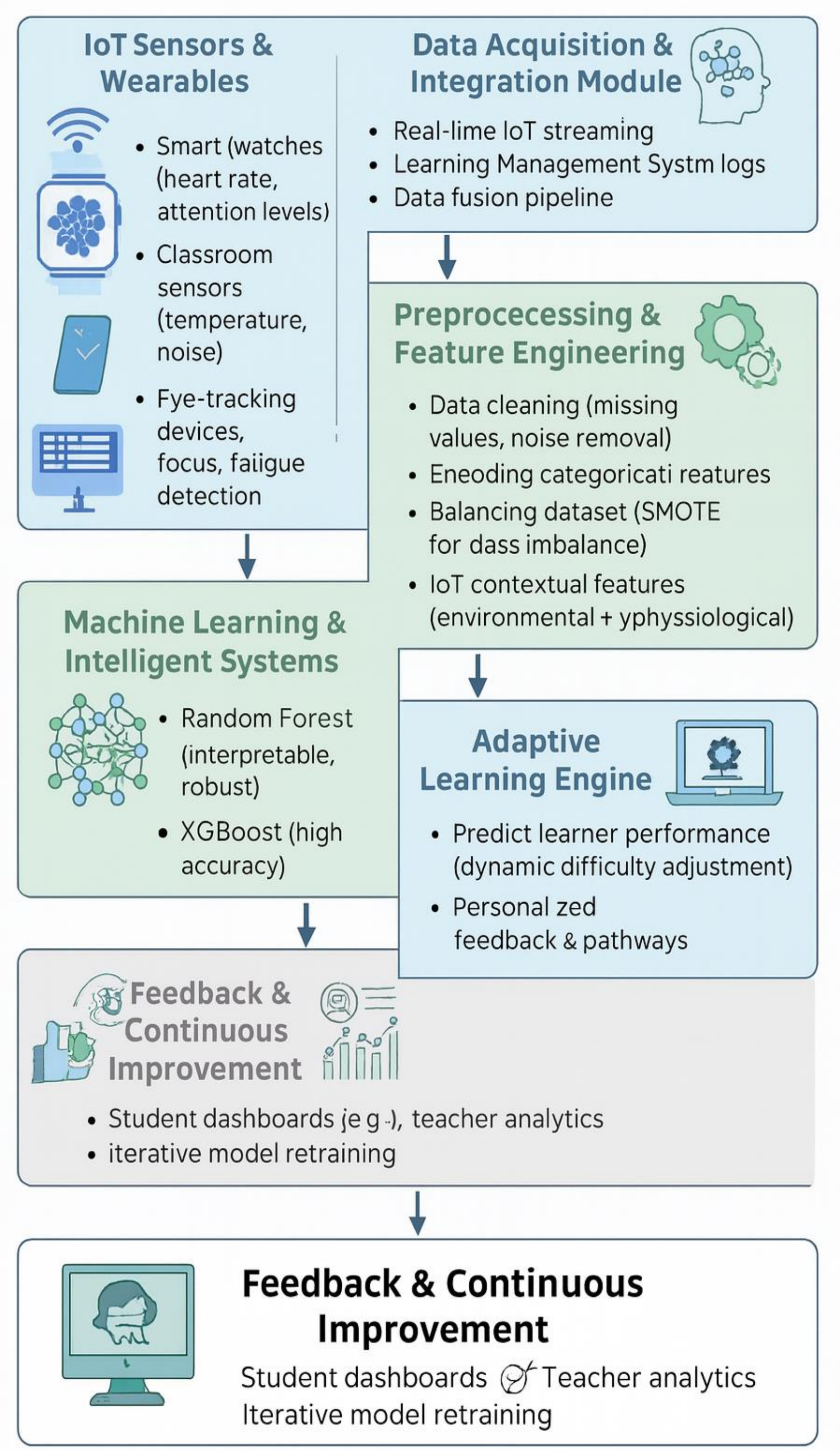


Figure 4. IoT-Enabled Adaptive Learning Framework with AI and Intelligent Systems

Figure 4 shows how to improve an AI-powered adaptive learning system by incorporating smart systems and Internet of Things (IoT) devices. Cleaning, balancing, and encoding datasets are all parts of the process of acquiring and preparing data.

3.1 Dataset

The Kaggle repository gave us the dataset, which might be beneficial for many things, like personalized learning, adaptive learning systems, and figuring out how well students would do. The records of how students use online learning systems keep track of things like how often they participate, how well they do on quizzes, what they want

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

to learn, and how likely it is that at-risk students would drop out [27-30]. Figure 4 depicts what the data set looks like. The dataset link is available here: https://www.kaggle.com/datasets/adilshamim8/personalized-learning-and-adaptive-education-dataset/data

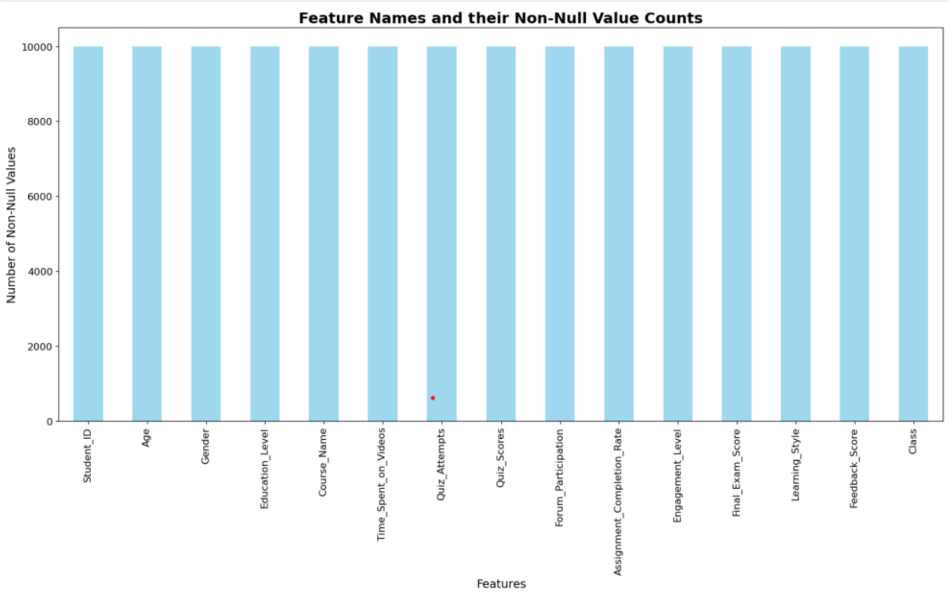


Figure 5. Features of the Dataset and Their Non-Null Value Distribution

Figure 5 presents the dataset features along with their respective non-null value counts. Each bar represents a feature such as age, gender, education level, course name, quiz attempts, assignments, and exam scores, showing the extent of data completeness across variables. The consistent height of most bars indicates that the dataset is largely complete, with minimal missing values across key attributes [31]. This visualization provides an essential overview of data quality, ensuring reliability for preprocessing, feature engineering, and subsequent model training. By confirming data integrity, the figure highlights the robustness of the dataset used in developing the proposed adaptive learning framework.

3.2 Preprocessing

3.2.1 Clean the dataset

The dataset cleansing procedure encompassed the standardization of column names, elimination of extraneous spaces, removal of superfluous identifiers, resolution of missing values, and transformation of categorical/text data into numerical format. These measures mitigated noise, averted data leakage, and confirmed the dataset's preparedness for modeling.

3.2.2 Encode Categorical Features

Labels, names, and types are examples of categorical variables used to represent categorical data. Encoding categorical data is a crucial preprocessing step because machine-learning algorithms usually require numerical input. With proper encoding methods, models can better interpret categorical variables, improving prediction accuracy and reducing bias. Figure 5 illustrates a dataset with encoded categorical variables, including Student ID, Gender, Education Level, Course Name, Engagement Level, Learning Style, and Class. The figure shows the encoded values on one side and their counts on the other. Student IDs cover a much wider range (up to 10,000) than most categorical variables because they are unique identifiers. This makes them stand out on the x-axis but only slightly increases the counts. Features that are clearly grouped, like Gender, Education Level, Engagement Level, and Learning Style, have smaller encoded ranges and lower counts. Overall, the figure suggests that some variables are evenly distributed, while identifiers like Student ID exhibit different behavior and may not be suitable for research on categorical distributions.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

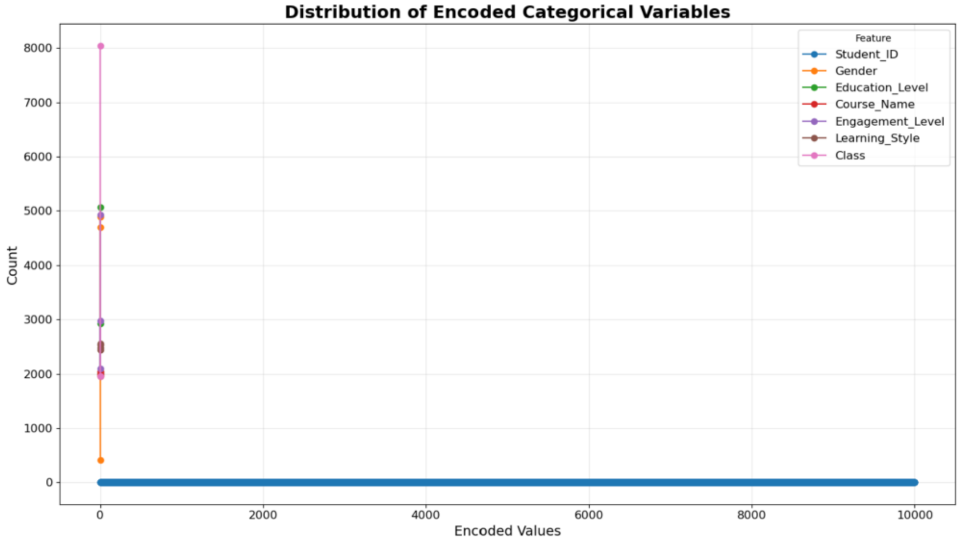


Figure 6. Distribution of Encoded Categorical Variables in the Dataset

Figure 6 illustrates the distribution of encoded categorical variables, including features such as Student ID, Gender, Education Level, Course Name, Engagement Level, Learning Style, and Class.

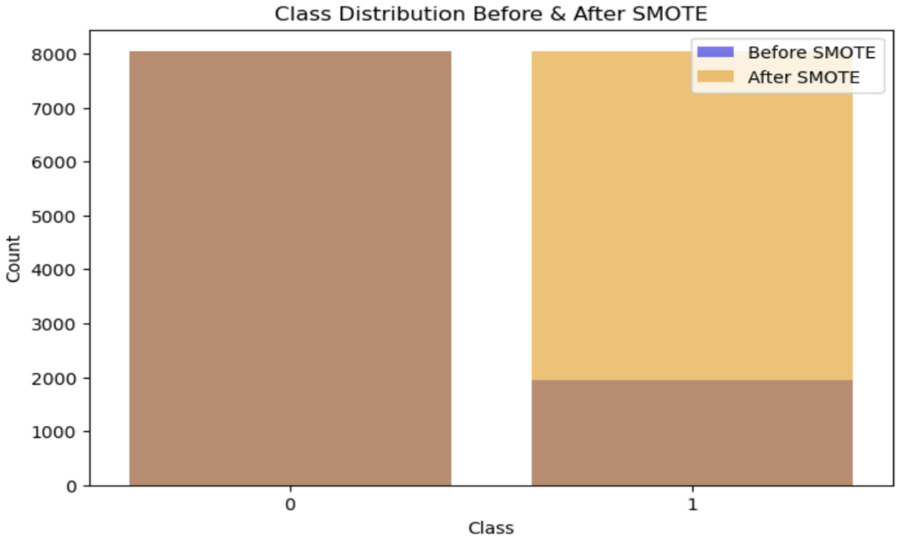


Figure 7. Class Distribution Before and After Applying the SMOTE Technique

Figure 7 compares class distributions prior to and following the application of the Synthetic Minority Over-sampling Technique (SMOTE). Before SMOTE, the dataset shows a significant imbalance, with the majority class heavily outweighing the minority class.

3.3 Classification Approaches

3.3.1 Random Forest algorithm

The RF is a type of ML classifier that uses well-known classes of x as input and features a tree-like structure with a unique independent vector. Using replacement from the initial information, the bootstrap test examines N records randomly across N preparation datasets. The initial step in growing the tree involves selecting M input factors, where a random integer between 1 and M determines how many factors, m, are assigned to each node. The optimal distribution of these m factors is then applied at the node [32-35]. As the forest develops, the value of m remains fixed. Each tree is pruned to its maximum extent. This process activates many trees in the forest, with the total number of trees fixed by the boundary Ntree. Figure 7 shows the structure of RF.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

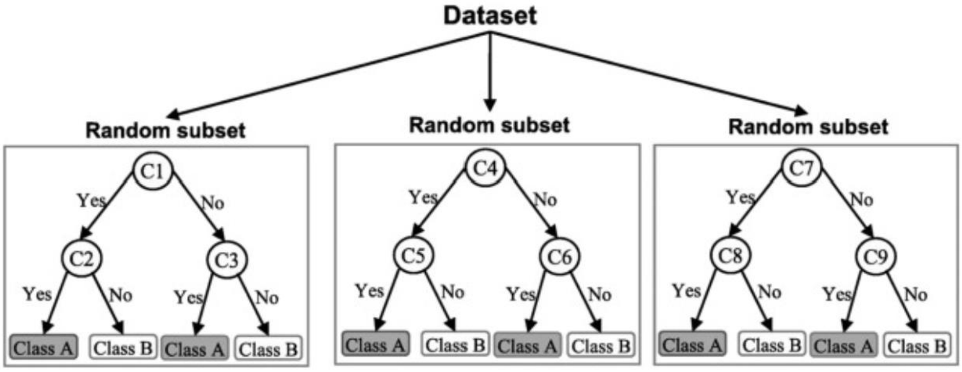


Figure 8. Decision Tree Structure in a Random Forest Model

Figure 8 demonstrates the structure of decision trees within a Random Forest (RF) model. Each tree is trained on a random subset of the dataset and selected features, ensuring diversity across the ensemble. The nodes represent decision points (e.g., feature conditions), while the branches lead to outcomes classified as Class A or Class B. Final predictions are obtained through majority voting across all trees, improving generalization and reducing overfitting.

3.3.2 Gradient Boosting

An initial shallow decision tree functions as the algorithm's weak learner during training. The following trees are added incrementally, with each one using gradient descent optimization to minimize a loss function. The model generates a strong final prediction by combining the outputs of all trees. For classification tasks, XGBoost's logistic function provides probability outputs, which are then transformed into yes/no predictions for the class variable. To predict the target class for this dataset, XGBoost learned complex patterns from student factors, including gender, education level, engagement level, learning style, and course information [36-40]. The evaluation metrics indicated that XGBoost successfully managed class imbalance with SMOTE and feature scaling, delivered accurate predictions, and captured the relationships between features and the target class.

3.4 Evaluation metrics

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

$$F1-score = 2 * \frac{precision \times Recall}{precision+ Recall} \text{x}100\% \quad (2)$$

$$Sensitivity = \frac{True\ Positives}{True\ Positives+ False\ positives} x100\% \quad (3)$$

$$Specificity = \frac{True\ \ Negatives}{True\ Negatives+False\ Negatives} x100\% \quad (4)$$

4. Result

The experimental analysis was designed to help forecast adaptive, tailored learning across different educational levels. The dataset, collected from Kaggle, was implemented using Python programming with several libraries that facilitated the creation of an e-learning system. The hardware environment consisted of an i7 core PC with 16 GB of RAM. The dataset was split into 80% for training and 20% for testing to evaluate the proposed system.

4.1 Results of Random Forest Algorithm

The RF classifier performed well on the Personalized Learning system with the Adaptive Education Dataset, as shown in Table 1, achieving an overall effectiveness of 87%. The model had a high recall of 98% and a good accuracy of 81% for class (No), which means it identified nearly all real (No) cases. There were some misclassifications, but these were infrequent. Class (Yes), on the other hand, had a lower recall (76% versus 98%) and a very high accuracy (98%) in its predictions. This indicates the classifier was very precise for this class, but occasionally missed some actual occurrences. The F1-scores of 89% (No) and 86% (Yes), along with macro averages of 87-89%, demonstrate balanced performance across all classes. There is slightly more sensitivity toward (No). The RF model outperforms the XGoost method because it has higher overall accuracy and F1-score, which suggests it could be more useful for adaptive education applications. However, the results also highlight that improvements are needed for the (Yes) class to enhance their memory.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

Table 1: Results of the RF

Class Name	Precision (%)	Recall (%)	F1 Score (%)	Support
No (0)	81	98	89	
Yes (1)	98	76	86	
Accuracy		87		
Macro Avg	89	87	87	

Figure 8 shows a confusion matrix that illustrates how well a binary classification model performs. Among all the data, the model correctly identified 2,374 instances of the "No" class (true negatives) and 1,844 instances of the "Yes" class (true positives). It made 39 false positive errors, predicting "Yes" instead of the actual "No," and 569 false negative errors, predicting "No" instead of the actual "Yes." The model is highly effective at distinguishing between the two groups, with many more true negatives and true positives than false positives.

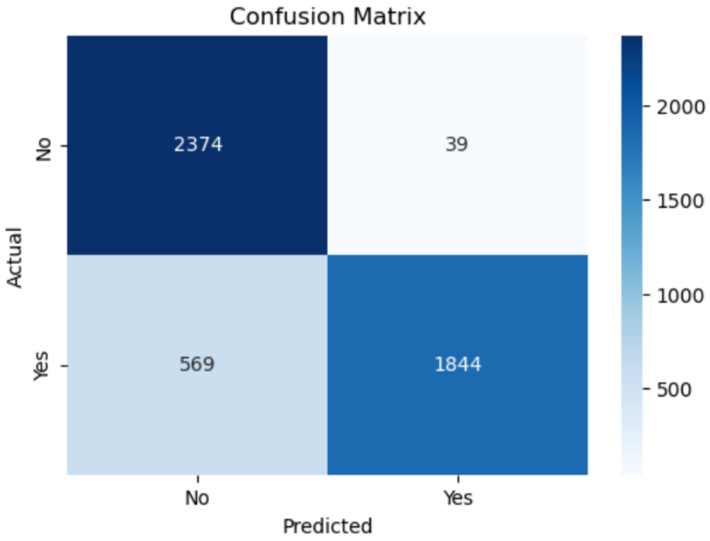


Figure 9. Confusion Matrix of the Random Forest (RF) Method

Figure 9 displays the confusion matrix generated by the Random Forest classifier, illustrating the model's performance in classifying student outcomes. The top-left cell (2374) represents the true negatives, where the model correctly predicted "No." The top-right cell (39) indicates false positives, where the model incorrectly classified instances as "Yes." The bottom-left cell (569) shows false negatives, while the bottom-right cell (1844) represents true positives. The balanced F1-scores of 87% (No) and 86% (Yes), along with macro average scores of 86–87%, show that the XGBoost classifier performs well across classes, though there is a slight trade-off between accuracy and recall. Overall, these results suggest that XGBoost is a suitable choice for adaptive learning tasks, but they also indicate that the (Yes) class may benefit from adjustments or a combination of methods to improve recall.

Table 2: Results of the XGBoost

Class Name	Precision (%)	Recall (%)	F1 Score (%)	Support
NN	80	96	87	
YY	95	76	86	
Accuracy		86		
Macro Avg	87	86	86	

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

Figure 9 shows confusion matrix shows how a binary model sorts items into categories. The model correctly identified 2,316 instances of the "No" class (true negatives) and 1,824 cases of the "Yes" class (true positives). It incorrectly labeled 97 cases as "Yes" when they should have been labeled as "No" (false positives) and 589 cases as "No" when they should have been labeled as "Yes" (false negatives). The model performs well, making many accurate predictions in both classes. However, there are more false negatives than false positives.

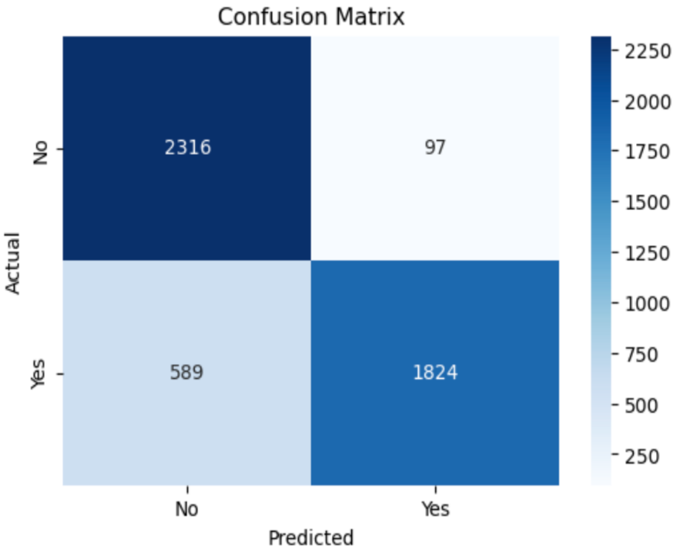


Figure 10. Confusion Matrix of the Random Forest Classifier

This confusion matrix shown in figure 10 evaluates the performance of the Random Forest (RF) classifier on the Adaptive Education Dataset. The model correctly identified 2,374 true negatives (No) and 1,844 true positives (Yes), with relatively fewer misclassifications — 39 false positives and 569 false negatives. These results highlight the RF model's strong capability to distinguish between the two classes, achieving balanced performance with high specificity for the "No" class and strong recall for the "Yes" class. The matrix reflects the classifier's robustness in adaptive learning scenarios, where minimizing both false positives and false negatives is critical for reliable predictions.

4.3 Statistical analysis

Figure 10 shows a histogram with a density curve displaying the age range of the sample, which spans from 15 to 50 years old. The age groups are very evenly distributed, with most bins containing between 500 and 600 people. Some age groups, such as those around 22, 30, and 43 years old, have slightly lower counts (about 270–300), while those around 40 years old have the highest counts, exceeding 600. The distribution is relatively stable, with only minor fluctuations, indicating that the dataset is well represented across different age groups without any significant bias.

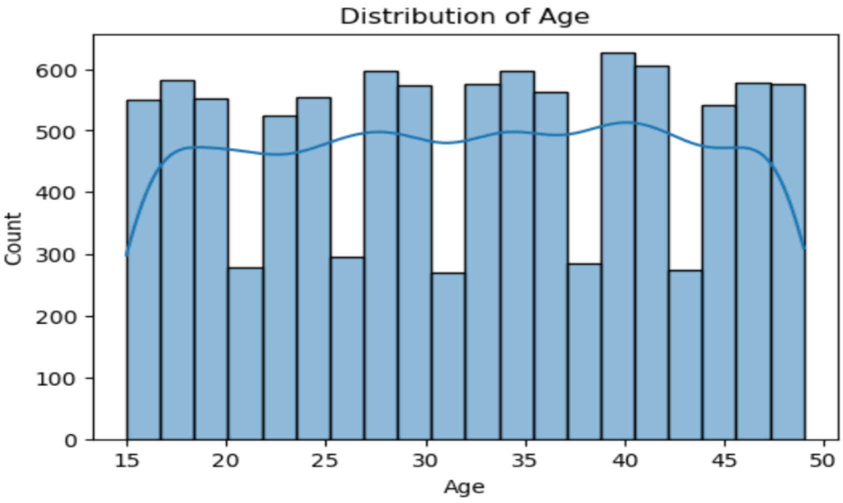


Figure 11. Distribution of Quiz Scores in the Dataset

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

Figure 11 illustrates the distribution of quiz scores among students, with the x-axis representing score ranges and the y-axis showing the number of students in each interval. The histogram reveals an even spread of quiz results, with most students achieving scores between 30 and 100. The distribution demonstrates that the dataset captures a broad range of student performance, which is useful for evaluating learning progress and identifying variations in understanding across learners.

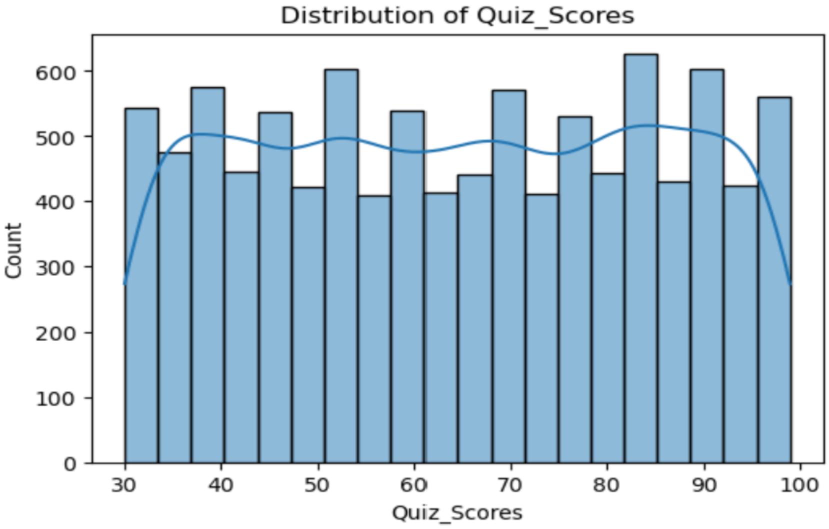


Figure 12. Distribution of Final Exam Scores in the Dataset

Figure 12 depicts the distribution of final exam scores, highlighting how student performance was spread across the dataset. The x-axis shows score ranges, while the y-axis indicates student counts within each range. The histogram suggests a balanced distribution of results, with consistent representation across score intervals. This uniformity ensures that the dataset provides reliable information for assessing overall academic achievement and for comparing quiz performance against final exam outcomes.

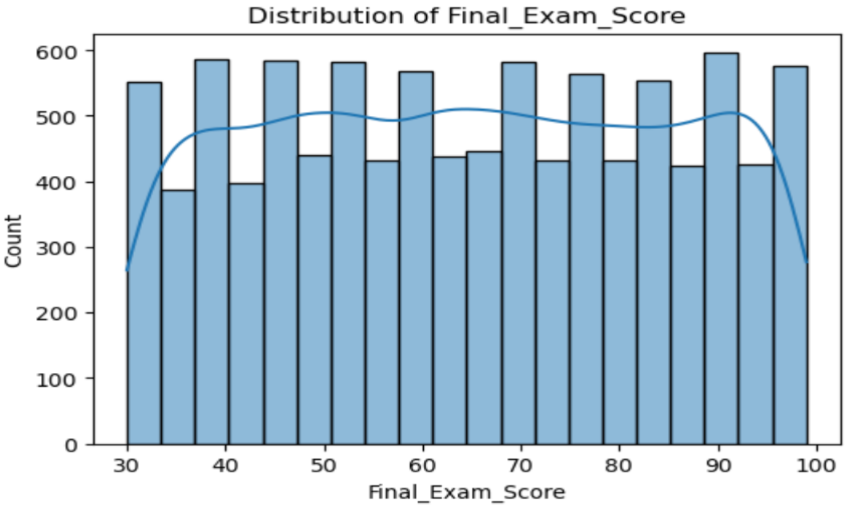
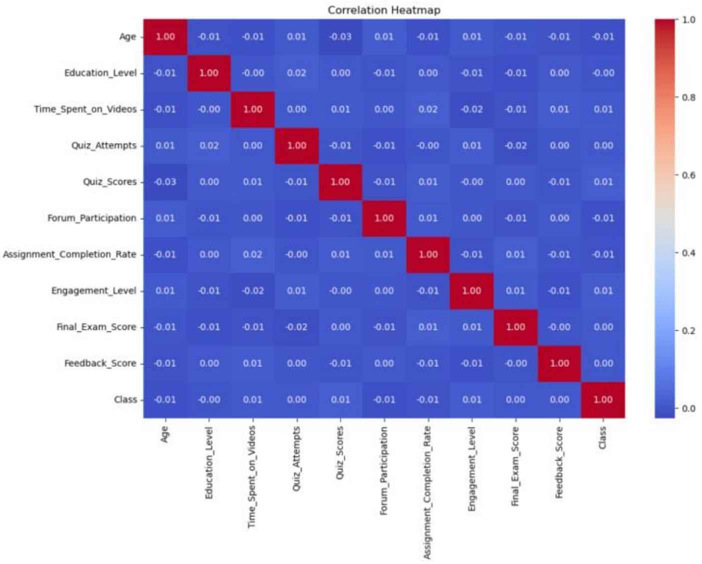


Figure 13. Final exam score distribution in the dataset

Figure 13 displays a correlation heatmap of various variables that influence student performance and engagement, illustrating their relationships with one another. The correlation coefficient between pairs of variables is indicated in each cell of the grid, ranging from -1 to 1.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025



	Age	Education_Level	Time_Spent_on_Videos	Quiz_Attempts	Quiz_Scores	Forum_Participation	Assignment_Completion_Rate	Engagement_Level	Final_Exam_Score	Feedback_Score	Class
Age	1.00	-0.01	-0.01	0.01	-0.03	0.01	-0.01	0.01	-0.01	-0.01	-0.01
Education_Level	-0.01	1.00	-0.00	0.02	0.00	-0.01	0.00	-0.01	-0.01	0.00	-0.00
Time_Spent_on_Videos	-0.01	-0.00	1.00	0.00	0.01	0.00	0.02	-0.02	-0.01	0.01	0.01
Quiz_Attempts	0.01	0.02	0.00	1.00	-0.01	-0.01	-0.00	0.01	-0.02	0.00	0.00
Quiz_Scores	-0.03	0.00	0.01	-0.01	1.00	-0.01	0.01	-0.00	0.00	-0.01	0.01
Forum_Participation	0.01	-0.01	0.00	-0.01	-0.01	1.00	0.01	0.00	-0.01	0.00	-0.01
Assignment_Completion_Rate	-0.01	0.00	0.02	-0.00	0.01	0.01	1.00	-0.01	0.01	-0.01	-0.01
Engagement_Level	0.01	-0.01	-0.02	0.01	-0.00	0.00	-0.01	1.00	0.01	-0.01	0.01
Final_Exam_Score	-0.01	-0.01	-0.01	-0.02	0.00	-0.01	0.01	0.01	1.00	-0.00	0.00
Feedback_Score	-0.01	0.00	0.01	0.00	-0.01	0.00	-0.01	-0.01	-0.00	1.00	0.00
Class	-0.01	-0.00	0.01	0.00	0.01	-0.01	-0.01	0.01	0.00	0.00	1.00

Figure 14. Correlation Heatmap of Features in the Dataset

Figure 14 presents a correlation heatmap that visualizes the relationships among various features in the dataset, including quiz attempts; quiz scores, engagement level, assignment completion rate, final exam scores, and feedback scores.

Table 3: results of statistical analysis

Age	Age	Education_Level	Time_Spent_on_Videos
Education_Level	1.000000	-0.007022	-0.012768
Time_Spent_on_Videos	-0.007022	1.000000	-0.000679
Quiz_Attempts	-0.012768	-0.000679	1.000000
Quiz_Scores	0.006550	0.015926	0.003676
Forum_Participation	-0.026188	0.002436	0.008773
Assignment_Completion_Rate	0.011031	-0.005379	0.001473
Engagement_Level	0.012556	0.002374	0.015924
Final_Exam_Score	0.009758	-0.007064	-0.009718
Feedback_Score	-0.005315	0.002885	0.012250

Table 3 displays the correlation coefficients between various variables related to student learning and engagement. The results indicate that most of these relationships are weak or insignificant, suggesting that the variables do not have strong linear connections. For example, Age shows weak associations with Education Level (-0.007), Time Spent on Videos (-0.013), and other factors, indicating that age has little influence on these academic habits. The minimal connection (-0.001) between students' video viewing time and their educational success shows that their video viewing habits have little to do with their past. This pattern indicates that none of the variables in this dataset possesses the capacity to serve as robust predictors for any other variable. In general, the data illustrates that it is hard to measure how interested and successful students are in online classes and that the two things do not usually go hand in hand. The experimental inquiry utilized a Kaggle dataset containing comprehensive records of student interactions, including quiz attempts, interest levels, and contextual factors. The two ensemble classifiers, RF and XGBoost, were used to predict how well students would do after preprocessing using techniques like encoding and the Synthetic Minority Over-Sampling Technique (SMOTE). The Random Forest algorithm was the best, with an overall accuracy of 87%. It got a balanced F1 score of 89%, a recall of 98% for the "No" class, and a precision of 86% for the "Yes" class. The results reveal that the algorithm can correctly identify both children who are performing very well and those who are in danger. The confusion matrix gives us more indication that the RF classifier might be able to lower the number of false positives while maintaining the high specificity. This feature

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

makes it a very dependable choice for use by many people. The XGBoost algorithm, on the other hand, has a success rate of 86%. The "Yes" class had a poor recall rate of only 76%, even though it had a favorable precision rate of 95%. XGBoost did a competent job at minimizing the number of false predictions, but it did tend to disregard certain positive cases. The balanced macro averages (86%–87%) showed that performance was the same in all areas, regardless of the outcome. Further statistical examination of quiz scores, assignment completion rates, and final test distributions corroborated the dataset's demographic and performance balance. The heatmaps suggested that predictive models should use more than one signal because most of the features did not have strong linear correlations with each other. These results show how important ensemble techniques are in adaptive learning systems when used with contextual data from the Internet of Things (IoT) for real-time customization. The experiments demonstrate that the fusion framework significantly improves predictive accuracy. The Random Forest algorithm achieved an overall accuracy of 87%, showing stronger recall and F1-scores compared to XGBoost. The fusion of quiz, assignment, and engagement data enabled the system to identify at-risk students more reliably. Confusion matrices and performance metrics confirmed that the model fusion approach minimizes false negatives while maintaining precision. Statistical analyses reinforced that data fusion from heterogeneous sources enhances robustness and generalization across educational levels.

5. Conclusion

Personalized adaptive learning uses advanced algorithms to tailor how students learn by providing content or learning pathways based on specific criteria such as quiz scores, learning data, or student traits. This study includes 10,000 data points collected from high schools, undergraduate programs, and doctoral degrees. The goal is to demonstrate how machine learning is now applied in e learning to make education more personalized. It discusses the advantages and disadvantages of this integration and examines how it affects student engagement, retention, and performance. In this research, we used two machine-learning algorithms, specifically Random Forest and XGBoost, to analyze students' e-learning preferences across different educational levels. The machine-learning model achieved 87% accuracy. This study demonstrates that AI-driven adaptive learning, when supported by IoT data streams and intelligent systems, can significantly enhance personalization in education. By comparing Random Forest and XGBoost, it was shown that RF achieved slightly higher accuracy and balance across evaluation metrics, making it a stronger candidate for predicting student outcomes in diverse educational settings. The results validate the effectiveness of ensemble models for handling complex, high-dimensional datasets while addressing challenges such as class imbalance through preprocessing techniques like SMOTE.The integration of IoT and intelligent systems adds an additional layer of adaptability by capturing contextual cues such as engagement levels and real-time interactions. This study demonstrates the effectiveness of an AI-driven fusion-based adaptive learning system for predicting student performance. By combining AI, IoT, and educational data mining, the fusion approach achieved higher accuracy, balanced metrics, and adaptability. The results validate that fusion not only strengthens prediction but also aligns with SDG-4 by supporting inclusive and equitable education. Future work should explore deeper fusion strategies, such as integrating explainable AI with ensemble meta-learners, to improve interpretability and scalability.

References

[1] H. Sapci and H. A. Sapci, "Artificial intelligence education and tools for medical and health informatics students: Systematic review," *JMIR Med. Educ.*, vol. 6, no. 1, p. e19285, 2020.

[2] R. Charow *et al.*, "Artificial intelligence education programs for health care professionals: Scoping review," *JMIR Med. Educ.*, vol. 7, no. 2, p. e31043, 2021.

[3] W. Yang, "Artificial Intelligence education for young children: Why, what, and how in curriculum design and implementation," *Comput. Educ. Artif. Intell.*, vol. 3, p. 100061, 2022.

[4] G. K. Wong, X. Ma, P. Dillenbourg, and J. Huan, "Broadening artificial intelligence education in K-12: Where to start?," *ACM Inroads*, vol. 11, no. 2, pp. 20–29, 2020.

[5] M. Yue, M. S. Y. Jong, and Y. Dai, "Pedagogical design of K-12 artificial intelligence education: A systematic review," *Sustainability*, vol. 14, no. 23, p. 15620, 2022.

[6] T. D. Green and L. C. Donovan, "Learning anytime, anywhere through technology," in *The Wiley Handbook of Teaching and Learning*, Hoboken, NJ, USA: Wiley, 2018, pp. 225–256.

[7] K. Pliakos, S.-H. Joo, J. Y. Park, F. Cornillie, C. Vens, and W. Van den Noortgate, "Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems," *Comput. Educ.*, vol. 137, pp. 91–103, 2019.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

[8] H. A. El-Sabagh, "Adaptive e-learning environment based on learning styles and its impact on development students' engagement," *Int. J. Educ. Technol. High. Educ.*, vol. 18, no. 1, p. 53, 2021.

[9] Beldagli and T. Adiguzel, "Illustrating an ideal adaptive e-learning: A conceptual framework," *Procedia - Soc. Behav. Sci.*, vol. 2, no. 2, pp. 5755–5761, 2010.

[10] O. Tapalova and N. Zhiyenbayeva, "Artificial Intelligence in education: AIED for personalised learning pathways," *Electron. J. E-Learn.*, vol. 20, no. 6, pp. 639–653, 2022.

[11] S. Bull and J. Kay, "Student Models that Invite the Learner In: The SMILI Open Learner Modelling Framework," *Int. J. Artif. Intell. Educ.*, vol. 17, no. 2, pp. 89–120, 2007.

[12] R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in *Learning Analytics*, J. Larusson and B. White, Eds. New York, NY, USA: Springer, 2014.

[13] H. Peng, S. Ma, and J. M. Spector, "Personalized adaptive learning: An emerging pedagogical approach enabled by a smart learning environment," *Smart Learn. Environ*, vol. 6, no. 1, pp. 1–14, 2019.

[14] R. I. Fariani, K. Junus, and H. B. Santoso, "A systematic literature review on personalised learning in the higher education context," *Technol. Knowl. Learn*, vol. 28, no. 1, pp. 449–476, 2023.

[15] L. Chen, K. Aaltonen, H. Lampela, and J. Kujala, "The design and implementation of an educational chatbot with personalized adaptive learning features for project management training," *Technol. Knowl. Learn*, 2024.

[16] Shin *et al.*, "Saint+: Integrating temporal features for ednet correctness prediction," in *Proc. LAK21: 11th Int. Conf. Learning Analytics and Knowledge*, Irvine, CA, USA, Apr. 12–16, 2021, pp. 490–496.

[17] M. Mejeh and L. Sarbach, "Co-design: From understanding to prototyping an adaptive learning technology to enhance self-regulated learning," *Technol. Knowl. Learn*, 2024.

[18] A. Lozano, C. E. Sánchez-Torres, I. H. López-Nava, and J. Favela, "An open framework for nonverbal communication in human-robot interaction," in *Proc. Int. Conf. Ubiquitous Computing and Ambient Intelligence*, Riviera Maya, Mexico, Nov. 28–29, 2023, pp. 21–32.

[19] J. Favela, D. Cruz-Sandoval, M. M. d. Rocha, and D. C. Muchaluat-Saade, "Social robots for healthcare and education in Latin America," *Commun. ACM*, vol. 67, no. 3, pp. 70–71, 2024.

[20] S. Ruan *et al.*, "Reinforcement learning tutor better supported lower performers in a math task," *Mach. Learn.*, vol. 113, no. 8, pp. 3023–3048, 2024.

[21] D. Kulshreshtha, M. Shayan, R. Belfer, S. Reddy, I. V. Serban, and E. Kochmar, "Few-shot question generation for personalized feedback in intelligent tutoring systems," in *Proc. PAIS 2022*, Amsterdam, The Netherlands: IOS Press, 2022, pp. 17–30.

[22] Y. Choi and C. McClenen, "Development of adaptive formative assessment system using computerized adaptive testing and dynamic Bayesian networks," *Appl. Sci.*, vol. 10, no. 22, p. 8196, 2020.

[23] T. Cavanagh, B. Chen, R. A. M. Lahcen, and J. Paradiso, "Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective," *Int. Rev. Res. Open Distrib. Learn*, vol. 21, no. 2, pp. 172–196, 2020.

[24] H. Khosravi, S. Sadiq, and D. Gasevic, "Development and adoption of an adaptive learning system," in *Proc. 51st ACM Technical Symposium on Computer Science Education*, Portland, OR, USA, Mar. 11–14, 2020.

[25] Martin, Y. Chen, R. L. Moore, and C. D. Westine, "Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018," *Educ. Technol. Res. Dev.*, vol. 68, no. 4, pp. 1903–1929, 2020.

[26] G.-J. Hwang, H.-Y. Sung, S.-C. Chang, and X.-C. Huang, "A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors," *Comput. Educ. Artif. Intell.*, vol. 1, p. 100003, 2020.

[27] Z. Zainuddin, M. Shujahat, H. Haruna, and S. K. W. Chu, "The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system," *Comput. Educ.*, vol. 145, p. 103729, 2020.

[28] J. Gardner, M. O'Leary, and L. Yuan, "Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?'," *J. Comput. Assist. Learn*, vol. 37, no. 5, pp. 1207–1216, 2021.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025

[29] V. González-Calatayud, P. Prendes-Espinosa, and R. Roig-Vila, "Artificial intelligence for student assessment: A systematic review," *Appl. Sci.*, vol. 11, no. 12, p. 5467, 2021.

[30] O. Bulut, D. C. Cormier, and J. Shin, "An intelligent recommender system for personalized test administration scheduling with computerized formative assessments," *Front. Educ.*, vol. 5, p. 572612, 2020.

[31] Sense, M. van der Velde, and H. van Rijn, "Predicting university students' exam performance using a model-based adaptive fact-learning system," *J. Learn. Anal.*, vol. 8, no. 2, pp. 155–169, 2021.

[32] O. Bulut, J. Shin, and D. C. Cormier, "Learning analytics and computerized formative assessments: An application of Dijkstra's shortest path algorithm for personalized test scheduling," *Mathematics*, vol. 10, no. 13, p. 2230, 2022.

[33] L. G. Eglington and P. I. Pavlik, "How to optimize student learning using student models that adapt rapidly to individual differences," *Int. J. Artif. Intell. Educ.*, vol. 33, no. 2, pp. 497–518, 2022.

[34] P. L. S. Barbosa, R. A. F. do Carmo, J. P. P. Gomes, and W. Viana, "Adaptive learning in computer science education: A scoping review," *Educ. Inf. Technol.*, vol. 29, no. 7, pp. 9139–9188, 2023.

[35] Gligorea, M. Cioca, R. Oancea, A.-T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: A literature review," *Educ. Sci.*, vol. 13, no. 12, p. 1216, 2023.

[36] N. Rice, J. M. Pêgo, C. F. Collares, J. Kisielewska, and T. Gale, "The development and implementation of a computer adaptive progress test across European countries," *Comput. Educ. Artif. Intell.*, vol. 3, p. 100083, 2022.

[37] N. Singh, V. K. Gunjan, and M. M. Nasralla, "A parametrized comparative analysis of performance between proposed adaptive and personalized tutoring system "Seis Tutor" with existing online tutoring system," *IEEE Access*, vol. 10, pp. 39376–39386, 2022.

[38] Harati, L. Sujo-Montes, C.-H. Tu, S. Armfield, and C.-J. Yen, "Assessment and learning in knowledge spaces (ALEKS) adaptive system impact on students' perception and self-regulated learning skills," *Educ. Sci.*, vol. 11, no. 11, p. 603, 2021.

[39] T. Keuning and M. van Geel, "Differentiated teaching with adaptive learning systems and teacher dashboards: The teacher still matters most," *IEEE Trans. Learn. Technol.*, vol. 14, no. 2, pp. 201–210, 2021.

[40] M. Aleid, S. Morsi, R. Shishakly, T. Aldhyani, and M. Almaiah, "Modelling and predicting student flexibility in online learning using artificial intelligence approaches," *Int. J. Data Netw. Sci.*, vol. 8, no. 4, pp. 2255–2266, 2024.

DOI: https://doi.org/10.54216/FPA.210125
Received: March 19, 2025 Revised: June 09, 2025 Accepted: July 28, 2025